



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2016

The Open Linguistics Working Group: Developing the Linguistic Linked Open Data Cloud

McCrae, John Philip ; Chiarcos, Christian ; Bond, Francis ; Cimiano, Philipp ; Declerck, Thierry ; de Melo, Gerard ; Gracia, Jorge ; Hellmann, Sebastian ; Klimek, Bettina ; Moran, Steven ; Osenova, Petya ; Pareja-Lora, Antonio ; Pool, Jonathan

Abstract: The Open Linguistics Working Group (OWLG) brings together researchers from various fields of linguistics, natural language processing, and information technology to present and discuss principles, case studies, and best practices for representing, publishing and linking linguistic data collections. A major outcome of our work is the Linguistic Linked Open Data (LLOD) cloud, an LOD (sub-)cloud of linguistic resources, which covers various linguistic databases, lexicons, corpora, terminologies, and metadata repositories. We present and summarize five years of progress on the development of the cloud and of advancements in open data in linguistics, and we describe recent community activities. The paper aims to serve as a guideline to introduce and involve researchers with the community and more generally with Linguistic Linked Open Data.

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-132353>

Conference or Workshop Item

Originally published at:

McCrae, John Philip; Chiarcos, Christian; Bond, Francis; Cimiano, Philipp; Declerck, Thierry; de Melo, Gerard; Gracia, Jorge; Hellmann, Sebastian; Klimek, Bettina; Moran, Steven; Osenova, Petya; Pareja-Lora, Antonio; Pool, Jonathan (2016). The Open Linguistics Working Group: Developing the Linguistic Linked Open Data Cloud. In: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016), Portorož, Slovenia, 23 May 2016 - 28 May 2016. European Language Resources Association (ELRA), 2435-2441.

The Open Linguistics Working Group: Developing the Linguistic Linked Open Data Cloud

John P. McCrae,¹ Christian Chiacaros,²
Francis Bond,³ Philipp Cimiano,⁴ Thierry Declerck,⁵ Gerard de Melo,⁶
Jorge Gracia,⁷ Sebastian Hellmann,⁸ Bettina Klimek,⁸ Steven Moran,⁹
Petya Osenova,¹⁰ Antonio Pareja-Lora,¹¹ Jonathan Pool¹²

¹Insight Centre for Data Analytics, National University of Ireland Galway, john@mccr.ae

²Goethe-University Frankfurt, Germany, chiacaros@informatik.uni-frankfurt.de

³Nanyang Technological University bond@ieee.org

⁴CIT-EC, Bielefeld University, cimiano@cit-ec.uni-bielefeld.de

⁵German Research Center for Artificial Intelligence, declerck@dfki.de

⁶IIS, Tsinghua University gdm@demelo.org

⁷Ontology Engineering Group, Universidad Politécnica de Madrid, jgracia@fi.upm.es

⁸InfAI, University of Leipzig, {hellmann, klimek}@informatik.uni-leipzig.de

⁹University of Zurich, steven.moran@uzh.ch

¹⁰IICT-BAS, petya@bultreebank.org

¹¹Universidad Complutense de Madrid / ATLAS (UNED), aplora@ucm.es

¹²Long Now Foundation, pool@panlex.org

Abstract

The Open Linguistics Working Group (OWLG) brings together researchers from various fields of linguistics, natural language processing, and information technology to present and discuss principles, case studies, and best practices for representing, publishing and linking linguistic data collections. A major outcome of our work is the Linguistic Linked Open Data (LLOD) cloud, an LOD (sub-)cloud of linguistic resources, which covers various linguistic databases, lexicons, corpora, terminologies, and metadata repositories. We present and summarize five years of progress on the development of the cloud and of advancements in open data in linguistics, and we describe recent community activities. The paper aims to serve as a guideline to introduce and involve researchers with the community and more generally with Linguistic Linked Open Data.

Keywords: Linked Data, language resources, community groups

1. The Open Linguistics Working Group

Linguistics, natural language processing, and related disciplines share a fundamental interest in language resources and their availability beyond individual research groups. This is necessary not only to fulfill fundamental principles of science (replicability), but also to facilitate subsequent re-use of resources created from public funding, e.g., as training data for novel tools, as a basis to increase the amount of data available for quantitative analyses, or as a component of innovative applications. The latter may include quite unforeseen uses, as in the case of the psycholinguistic resource WordNet (Fellbaum, 1998) turning into a significant component in numerous information technology systems.

Publishing language resources under open licenses, to facilitate exchange of knowledge and information across boundaries between disciplines as well as between academia and the IT business, has thus been an area of increasing interest in academic circles, including applied linguistics, lexicography, computational linguistics, and information technology. Interested individuals began to organize themselves in the context of the Open Knowledge Foundation (OKFN)¹, a community-based non-profit organization aiming to promote open data. In 2010, we established the **Open Linguistics Working Group (OWLG)**² of the Open Knowledge

Foundation as an interdisciplinary network open to anyone interested in publishing and using language resources, or in open licenses. The OWLG facilitates information exchange through a mailing list, regular meetings, joint publications, community projects, and interdisciplinary workshops, including the Linked Data in Linguistics workshop series with recent editions in Reykjavik and Beijing, and the next event in this series co-located with this conference (LREC 2016) in Portorož, Slovenia.

In parallel to forming the OWLG, we have seen a rising interest in using Semantic Web standards to represent web-accessible, but distributed and heterogeneous language resources in a uniform and interoperable way, and as a means to facilitate the access of openly available language resources. As these two trends in the field converged, the OWLG not only spearheaded the creation and collection of open linguistic data, but also initiated the creation of the **Linguistic Linked Open Data (LLOD) Cloud** (Sect. 2.), our most influential community project.

Subsequently, we have seen not only a number of approaches to provide linguistic data as linked data, but also the emergence of additional initiatives that aim to interconnect these resources (Sect. 3.). The LLOD cloud is hence being developed in collaboration with several W3C Community Groups and with European projects, especially Lider³, which was a community support action for linguis-

¹<http://okfn.org/>

²<http://linguistics.okfn.org>

³<http://www.lider-project.eu>

tic linked data, and also LOD2⁴ and QTLeap.⁵ These efforts have led to a growth of about 357% since the first instantiation of the cloud (28 linked resources in February 2012, 128 in February 2016).

2. The LLOD Cloud

The Linguistic Linked Open Data Cloud is the most prominent community project of the OWLG. It was established to measure and visualize the adoption of linked and open data within the linguistics community. Since its first conceptualization in 2011 and its first materialization in 2012, considerable work has gone into improving the definition and infrastructure that supports the cloud, with the result that, since Spring 2015, the cloud is generated on a monthly basis, as shown in Figure 1. In particular, we have further refined the criteria for inclusion and the methods for tracking resources.

2.1. Linguistically Relevant Resources

An important criterion is that a dataset must be *linguistically relevant* in that it provides or describes language data that can be used for the purpose of linguistic research or natural language processing. In particular, we define the following kinds of resources:

1. **Linguistic resources in a strict sense** are resources that were intentionally created for the purpose of linguistic research or natural language processing, and which contain linguistic classifications, annotations, or analyses or have been used to provide such information about language data.
2. **Other linguistically relevant resources** include all other resources used for linguistic research or natural language processing, but not necessarily created for this purpose, e.g., large collections of texts such as news articles, terminological or encyclopedic and general-purpose knowledge bases such as DBpedia (Bizer et al., 2009), or metadata collections.

2.2. Infrastructure and Metadata

The OWLG provides guidelines to data publishers on how to include their resources in the LLOD cloud.⁶ The cloud diagram is currently generated from metadata maintained at DataHub⁷ and hence contains only resources described in DataHub. An alternative metadata repository specialized for linguistic resources is under development: Linghub (McCrae et al., 2015a).⁸ It aims to provide a search engine and index for linguistic resources and attempts to harmonize metadata from a number of different sources, including Metashare (Federmann et al., 2012), CLARIN VLO (Van Uytvanck et al., 2012), DataHub and LRE Map (Calzolari et al., 2012). It will soon replace DataHub in the generation of the cloud diagram. LingHub,

⁴<http://lod2.eu/>

⁵<http://qt leap.eu/>

⁶http://wiki.okfn.org/Working_Groups/Linguistics/How_to_contribute

⁷<http://datahub.io>

⁸<http://linghub.org>

Date	Datasets	Links
February 2012	28	41
September 2013	53	78
November 2014	103	167
May 2015	126	203
February 2016	128	209

Table 1: Growth of the LLOD cloud over time

being an indexing and search service, only harvests, processes and indexes metadata from external repositories but does not support the direct upload or submission of language resources, which should be done via its component repositories.

We classify LLOD resources into three broad groups:

Corpora (blue in Fig. 1) are collections of language data, e.g., examples, text fragments, or entire discourses.

Lexical-conceptual resources (green in Fig. 1) focus on the general meaning of words and the structure of semantic concepts.

Metadata (red in Fig. 1) includes resources providing information about language and language resources, i.e., typological databases (collections of features and inventories of individual languages, e.g., from linguistic typology), linguistic terminology repositories (e.g., grammatical categories or language identifiers), and metadata about language resources (linguistic resource metadata repositories, including bibliographical data).

Among LLOD data sets, we encourage the use of **open** licenses. As defined by the Open Definition, *open* refers to “[any] piece of content or data [that] is open if anyone is free to use, reuse, and redistribute it – subject only, at most, to the requirement to attribute and share-alike.”⁹ At the moment, this condition is monitored but not strictly enforced as a criterion for inclusion in the LLOD cloud. This is in part due to the lack of information about the licensing of the resources and ongoing discussions within the group about the use of non-commercial licenses. However, we expect to reach a consensus within the next few months.

2.3. Extracting the LLOD Cloud

The LLOD cloud is extracted on the basis of the metadata in Datahub¹⁰. These resources are collected directly by means of the API and validated using the steps described below. Then a D3.js¹¹ script is used to generate the image and update the site, which is carried out on a monthly basis. All the diagrams are available at <http://linguistic-lod.org> and some statistics are given in Table 1, showing a continuous growth of the cloud in the past years.

2.4. Validation

The first draft of the LLOD diagram, presented at LREC 2012 (Chiarcos et al., 2011), still included many resources whose providers had at the time merely promised to provide linked open data. The criteria for inclusion have

⁹<http://opendefinition.org>

¹⁰<http://datahub.io>

¹¹<https://d3js.org>



Figure 1: Linguistic Linked Open Data cloud as of October 2015

subsequently been strengthened by requiring availability of a resource and its links (thus manifesting the first actual LLOD diagram rather than a draft, since September 2012), metadata quality, etc. Since early 2015, we introduced increasing *automatic* verification routines for the metadata provided by the resource providers. In order to bring resources into the linked data cloud we rely on the metadata recorded in Datahub. In particular, we attempt to find resources by looking for specific groups and tags that are associated with linguistic resources. We then check that the resource's metadata includes some link to some other resource in the LLOD cloud, and we hope to automatically detect the links in the immediate future by building on the work of the LODvader project.¹² We then check that the resource is available by attempting to download it and discarding all resources that are no longer available. We have attempted to notify the authors of resources that no longer meet the criteria for inclusion in the cloud. However, our experience has been that this did not motivate many authors to update their resources.

2.5. Vocabularies

The Linguistic Linked Open Data Cloud has grown significantly in the last few years and most notably, unlike the non-linguistic LOD Cloud, is not centered around one nu-

cleus but instead has used many different vocabularies and datasets to link to. Among these are BabelNet (Ehrmann et al., 2014), LexInfo (Cimiano et al., 2011), and Lexvo (de Melo, 2015). In addition, a number of new vocabularies have emerged including the OntoLex model,¹³ the NLP Interchange format NIF (Hellmann et al., 2013), the WordNet Interlingual Index (Sect. 5.2.), and the FrameBase schema (Rouces et al., 2015a) (Sect. 5.3.).

These vocabularies have increased the power of linked data to represent the complete spectrum of language resources and show that new resources can be created that use the power of linked data to link across different types of languages resources, such as terminologies and dictionaries (Siemoneit et al., 2015) and corpora and dictionaries (McGovern et al., 2015).

3. Other Community Group Efforts

OWLWG members have been very active in promoting the development and adoption of linguistic linked data, which had an effect not only in the growth of the LLOD cloud but in the development of representation models, guidelines, and best practices. These activities have been developed in the context of a number of W3C groups and projects, as it is detailed in the rest of this section.

¹²<http://lodvader.aksw.org/>

¹³<http://cimiano.github.io/ontolex/specification.html>

3.1. OntoLex

The **Ontology-Lexica Community (OntoLex) Group**¹⁴ was founded in September 2011 as a W3C Community Group. It aims to produce specifications for a lexicon-ontology model that can be used to provide rich linguistic grounding for domain ontologies. Rich linguistic grounding includes the representation of morphological and syntactic properties of lexical entries as well as the syntax-semantics interface, i.e., the meaning of these lexical entries with respect to a given ontology. An important issue herein will be to clarify how extant lexical and language resources can be leveraged and reused for this purpose. As a byproduct of this work on specifying a lexicon-ontology model, we are establishing a network of lexical and terminological resources that are linked according to the Linked Data principles, forming a large network of lexico-syntactic knowledge.

3.2. LIDER, BPMLOD and LD4LT

The LIDER project was a support action funded by the FP7 European program aimed to exploit and build upon multilingual and linguistic linked data for content analytics by establishing a strategy for progressing from existing industry practices and technological capabilities to the vision of the LLOD. The project built a global community of stakeholders in industry, research and standards, interested in the use of LLOD for multilingual, cross-media content analytics. Upon the conclusion of the project, we understand that OWLG can play an important role in the continuation of the community. The main outcomes of the project have been the development of a *reference architecture* (Brümmer et al., 2015) and a *roadmap* (Cimiano et al., 2015) for LOD-based multilingual, cross-media content analytics in enterprises.

Also with the support of the LIDER project, the W3C Best Practices for Multilingual Linked Open Data (BPMLOD) community group¹⁵ have developed a set of guidelines and best practices for integrating language and media resources into the LOD cloud, as well as generating and exploiting LOD-based language and media resources for content analytics. This has constituted an important step towards the dissemination and adoption of LLOD. The referred guidelines are:

1. Linguistic Linked Data Generation: Multilingual Dictionaries (BabelNet)
2. Linguistic Linked Data Generation: Bilingual Dictionaries
3. Linguistic Linked Data Generation: Multilingual Terminologies (TBX)
4. Developing NIF-based NLP Web Services
5. LLD Exploitation
6. Linguistic Linked Data Generation: WordNets
7. Linked Data corpus creation using NIF

8. LLOD aware services

Further, LIDER has produced eight practical reference cards that provide easy-to-follow recipes for the publication of linguistic resources as linked data¹⁶:

1. How to publish Linguistic Linked Data
2. Language Resource Licensing
3. Inclusion in the LLOD Cloud
4. Data ID
5. Discovering Language Resources with LingHub
6. NIF corpus
7. How to represent crosslingual links
8. Documenting a language resource in Datahub

Most of these guidelines and reference cards have been developed by members of OWLG and feedback have been gathered through the BPMLOD and OWLG groups. Finally, we shall mention the W3C Linked Data for Language Technologies (LD4LT) community group¹⁷. Its activities, complementary to those at OWLG, have been focused on gathering use cases and requirements from industry for linguistic linked data based content analytics. Also, it served as forum for discussions such as the convergence of broadly accepted metadata schemes into a unified model for describing language resources, which resulted in the OWL model that LingHub adopted (McCrae et al., 2015b). Due to the common goals and interest of both LD4LT and OWLG, the possibility of creating joint mailing lists and occasional joint calls is under consideration.

4. Community Events

Since the foundation of the OWLG in 2010, various events have taken place, which went along with its main goals of promoting the creation of open data in linguistics and intensifying the communication between researchers from different communities that use, distribute, or maintain open linguistic data.

A very well received event is the OWLG-organized workshop series on **Linked Data in Linguistics (LDL)**, which was established in 2012 and attracts an interdisciplinary and international community on a yearly basis. A significant outcome of the recent editions in Reykjavik (2014) and Beijing (2015) has been to facilitate the publishing of LLOD resources and vocabularies, applications and use cases. As a result, 10 papers and at least 3 posters, including this year's edition of LDL in co-location with LREC 2016, have contributed to the community's topics of interest.

In addition, efforts have gone into supporting and educating junior researchers, e.g. by dedicating entire summer schools to Linguistic Linked Open Data (**12th EUROLAN summer school, 2015**)¹⁸ and by giving introductory LLOD

¹⁴<http://www.w3.org/community/ontolex>

¹⁵<http://www.w3.org/community/bpmlod/>

¹⁶<http://www.lider-project.eu/guidelines>

¹⁷<https://www.w3.org/community/ld4lt/>

¹⁸<http://eurolan.info.uaic.ro/2015>

courses at summer schools (**ESSLLI-2015**)¹⁹. Also, being the first event of this kind, the **Summer Datathon on Linguistic Linked Open Data (SD-LLOD-15)**²⁰ directly promoted the (re-)use and contribution to the LLOD cloud by training people from industry and academia, providing practical knowledge in the field of Linked Data applied to linguistics, as well as enabling participants to migrate their own (or other's) linguistic data and publish them as Linked Data on the Web.

The underlying Linked Data formats of the linguistic resources requires solid knowledge from Semantic Web and NLP experts in order to optimally exploit the LLOD cloud datasets according to the various needs of the different OWLG community members. Therefore, events such as the workshop series on the **Multilingual Semantic Web (MSW)**²¹ and **Natural Language Processing and Linked Open Data (NLP&LOD)**²² focused on the technical side of LLOD cloud content. Questions such as how recent advances in the area of Linked Open Data and NLP can be used synergistically have been explored by working on topics such as enhancing NLP applications with LOD, information extraction from LOD using NLP techniques, manipulating LOD with NLP techniques, LOD as a corpus or mapping LOD to common sense ontologies and language data.

Given the central focus on language data within the OWLG, events that aim at increasing the involvement of linguists are of great importance. Occasions such as the **Association for Linguistic Typology 10th Biennial Conference**²³ and the **LLOD workshop at the Summer Institute of the Linguistic Society of America**²⁴ have been taken as opportunities to not only present the Linked Data standards as a new method for language resource representation to linguists but also to learn from their longstanding expertise and concomitant challenges of language data compilation, comparison and re-use for linguistic research.

Due to the openness of the language resources provided in the LLOD cloud, practitioners, industry and infrastructure providers operating across language barriers are increasingly interested in the possibilities offered by the multilingual Linked Data resources. Hence, OWLG community members have been engaged in events such as the **Multilingual Linked Open Data for Enterprises (MLODE 2014)**²⁵ in order to discuss how to channel feedback from industry to open source and academic communities. Industry representatives, researchers and engineers examined industrial use cases and the building of LOD-aware NLP services.

Finally, a special issue of the Semantic Web Journal on Multilingual Linked Open Data was published in

2015 (McCrae et al., 2015c)²⁶, including the description of 10 papers, out of which 7 were dataset descriptions, demonstrating the deep scientific impact of the topic.

5. Use Cases

5.1. Multilingual Processing of Linked Data for the Legal Domain

Within the EUCases project (<http://eucases.eu/start/>), an EUCases Legal Linked Open Dataset (EUCases-LLOD) was created. First, the legal data was processed via a multilingual pipeline (Bulgarian, Italian, German, French and English) for identifying named entities and EuroVoc concepts. Then, the XML documents were transformed to RDF. The dataset is linked to EuroVoc and the Syllabus ontology, since they are used as domain specific ontologies. Additionally, other supporting ontologies have been added, such as GeoNames for the named entities; PROTON as an upper ontology; SKOS as a mapper between ontologies and terminological lexicons; Dublin Core as a metadata ontology. Also, for the purposes of search, Web Interface Querying EUCases Linking Platform was designed. For its Web Interface, the EUCases Linking Platform relies on a customized version of the GraphDB Workbench²⁷, developed by Ontotext AD.

5.2. Wordnet Interlingual Index (ILI)

A recent development (Vossen et al., 2016; Bond et al., 2016) has been the adoption of LLOD technology by the wordnet community, with a new plan that uses LLOD as the basic mechanism for the creation of links between wordnets in different languages. This Collaborative InterLingual Index enables wordnets to share and link their resources for concepts lexicalized in any of the group's languages. This was supported directly by a workshop at the 2016 Global WordNet Conference and will lead to the adoption of LLOD technology by a new community. In addition, the open multilingual wordnet (Bond et al., 2014) provides all open wordnets for download using the lemon (McCrae et al., 2012) and RDF standard.

5.3. FrameBase

FrameBase (Rouces et al., 2015a) is a new large-scale vocabulary, based on FrameNet and WordNet, that uses the linguistic notion of semantic frames for general-purpose knowledge representation. It can thus serve as a bridge between the linguistic knowledge in the LLOD cloud and the general LOD cloud. Knowledge from different sources such as YAGO and Freebase can be mapped into this schema, even if they use very heterogeneous forms of representations. For instance, YAGO has an *isMarriedTo* relation and uses a form of reification to describe its properties, while other sources model a marriage as an instance. FrameBase was used in the European Union FP7 project ePOOLICE as part of an integrated knowledge repository to represent knowledge related to organized crime obtained

¹⁹<http://esslli2015.org>

²⁰<http://datathon.lider-project.eu>

²¹<http://msw2.deri.ie>

²²<http://bultreebank.org/NLP&LOD>

²³https://www.eva.mpg.de/lingua/conference/2013_ALT10

²⁴<http://quijote.fdi.ucm.es:8084/LLOD-LSASummerWorkshop2015/Home.html>

²⁵<https://mlode2014.nlp2rdf.org>

²⁶<http://www.semantic-web-journal.net/blog/call-multilingual-linked-open-data-mlod-2012/-data-post-proceedings>

²⁷<http://graphdb.eucases.eu/graphdb>

from several conceptual graphs (Rouces et al., 2015b). Further details are available on the FrameBase website²⁸.

6. Conclusion

The activity and impact made by the use of linked and open data have been significant in the last few years, as shown by the increasing growth of the LLOD cloud and the wealth of events and groups that have developed within the last two years. However, as the community has grown larger, it has also become harder to manage, and new tools are needed to motivate the continuing adoption of the paradigm. We believe that with the strong basis of our existing community these will easily be met and lead to further revolutionary change in the field of language resources.

7. Acknowledgements

This work has been supported by the LIDER FP7 European project (ref. 610782); the Spanish Ministry of Economy and Competitiveness through the project 4V (TIN2013-46238-C4-2-R), the Excellence Network ReTeLe (TIN2015-68955-REDT), the Juan de la Cierva program; the Science Foundation Ireland under Grant Number SFI/12/RC/2289 (Insight); the FREME FP7 European project (ref.GA-644771); by the QTLep FP7 EU project under grant agreement number 610516; by the Singapore MOE Tier 2 grant (ARC41/13); by the National Science Foundation of the USA (NSF, Award No. 1463196) and by China 973 Program Grants 2011CBA00300, 2011CBA00301, and NSFC Grants 61033001, 61361136003, 61550110504.

8. References

- Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., and Hellmann, S. (2009). DBpedia - a crystallization point for the web of data. *Web Semantics: Science, Services and Agents on the World Wide Web*, 7(3):154–165, September.
- Bond, F., Fellbaum, C., Hsieh, S.-K., Huang, C.-R., Pease, A., and Vossen, P. (2014). A multilingual lexico-semantic database and ontology. In Paul Buitelaar et al., editors, *Towards the Multilingual Semantic Web*, pages 243–258. Springer.
- Bond, F., Vossen, P., McCrae, J. P., and Fellbaum, C. (2016). CILI: the Collaborative Interlingual Index. In *Proc. of the Eighth Global WordNet Conference (GWC 2016)*, pages 50–57.
- Brümmer, M., Hellmann, S., Ackermann, M., Koidl, K., Lewis, D., Cimiano, P., Hartung, M., McCrae, J., Unger, C., Rodríguez-Doncel, V., Gómez-Pérez, A., Gracia, J., Flati, T., Navigli, R., Moro, A., and Buitelaar, P. (2015). Linguistic linked data reference architecture – phase II. Technical report, LIDER project deliverable, October.
- Calzolari, N., Del Gratta, R., Francopoulo, G., Mariani, J., Rubino, F., Russo, I., and Soria, C. (2012). The LRE Map. Harmonising community descriptions of resources. In *Proc. of the 8th International Conference on Language Resources and Evaluation*, pages 1084–1089.
- Chiarcos, C., Hellmann, S., and Nordhoff, S. (2011). Towards a linguistic linked open data cloud: The Open Linguistics Working Group. *TAL*, 52(3):245–275.
- Cimiano, P., Buitelaar, P., McCrae, J., and Sintek, M. (2011). Lexinfo: A declarative model for the lexicon-ontology interface. *Web Semantics: Science, Services and Agents on the World Wide Web*, 9(1):29–51.
- Cimiano, P., Hartung, M., Gómez-Pérez, A., de Cea, G. A., Montiel-Ponsoda, E., Rodríguez-Doncel, V., Lewis, D., Buitelaar, P., Navigli, R., and Flati, T. (2015). Roadmap for the use of linguistic linked data for content analytics – phase II. Technical report, LIDER project deliverable, October.
- de Melo, G. (2015). Lexvo.org: Language-related information for the linguistic linked data cloud. *Semantic Web*, 6(4):393–400.
- Ehrmann, M., Cecconi, F., Vannella, D., McCrae, J. P., Cimiano, P., and Navigli, R. (2014). Representing multilingual data as linked data: the case of BabelNet 2.0. In *Proc. of 9th Language Resources and Evaluation Conference*, pages 401–408.
- Federmann, C., Giannopoulou, I., Girardi, C., Hamon, O., Mavroeidis, D., Minutoli, S., and Schröder, M. (2012). META-SHARE v2: An open network of repositories for language resources including data and tools. In *Proc. of the 8th International Conference on Language Resources and Evaluation*, pages 3300–3303.
- Christine Fellbaum, editor. (1998). *WordNet: An Electronic Lexical Database*. MIT Press.
- Hellmann, S., Lehmann, J., Auer, S., and Brümmer, M. (2013). Integrating NLP using linked data. In *The Semantic Web – ISWC 2013*, pages 98–113. Springer.
- McCrae, J., de Cea, G. A., Buitelaar, P., Cimiano, P., Declerck, T., Gómez-Pérez, A., Gracia, J., Hollink, L., Montiel-Ponsoda, E., Spohr, D., and Wunner, T. (2012). Interchanging lexical resources on the Semantic Web. *Language Resources and Evaluation*, 46(6):701–709.
- McCrae, J. P., Cimiano, P., Doncel, V. R., Vila-Suero, D., Gracia, J., Matteis, L., Navigli, R., Abele, A., Vulcu, G., Buitelaar, P., et al. (2015a). Reconciling heterogeneous descriptions of language resources. In *Fourth Workshop on Linked Data in Linguistics*, pages 39–74.
- McCrae, J. P., Labropoulou, P., Gracia, J., Villegas, M., Doncel, V. R., and Cimiano, P. (2015b). One ontology to bind them all: The META-SHARE OWL ontology for the interoperability of linguistic datasets on the web. In *Proc. of 12th Extended Semantic Web Conference (ESWC 2015) Satellite Events, Portorož, Slovenia*, volume 9341, pages 271–282, June.
- McCrae, J. P., Moran, S., Hellmann, S., and Brümmer, M. (2015c). Multilingual linked data (editorial). *Semantic Web*, 6(4):315–317.
- McGovern, A., O’Connor, A., and Wade, V. (2015). From DBpedia and WordNet hierarchies to LinkedIn and Twitter. In *Proc. of the 4th Workshop on Linked Data in Linguistics: Resources and Applications (LDL-2015)*, pages 1–10.
- Rouces, J., de Melo, G., and Hose, K. (2015a). Framebase:

²⁸<http://www.framebase.org>

- Representing n-ary relations using semantic frames. In *Proc. of ESWC 2015*.
- Rouces, J., de Melo, G., and Hose, K. (2015b). Representing specialized events with FrameBase. In *Proceedings of ESWC 2015 DeRiVE Workshop*.
- Siemoneit, B., McCrae, J. P., and Cimiano, P. (2015). Linking Four Heterogeneous Language Resources as Linked Data. In *Proc. of the 4th Workshop on Linked Data in Linguistics: Resources and Applications (LDL-2015)*, pages 59–63.
- Van Uytvanck, D., Stehouwer, H., and Lampen, L. (2012). Semantic metadata mapping in practice: the virtual language observatory. In *Proc. of the 8th International Conference on Language Resources and Evaluation*, pages 1029–1034.
- Vossen, P., Bond, F., and McCrae, J. P. (2016). Toward a truly multilingual Global Wordnet Grid. In *Proc. of the Eighth Global WordNet Conference (GWC 2016)*, pages 419–427.